Subjective and Objective Measures of Symptoms, Function, and Outcome in Patients With Degenerative Spine Disease

Maldaner, Nicolai ; Stienen, Martin Nikolaus

DOI: https://doi.org/10.1002/acr.24210

Arthritis Care & Research
Vol. 72, No. S10, October 2020, pp 183–199
DOI 10.1002/acr.24210

American College of Rheumatology
Empowering Rheumatology Professionals

Subjective and Objective Measures of Symptoms, Function, and Outcome in Patients With Degenerative Spine Disease

Nicolai Maldaner[1] and Martin Nikolaus Stienen[2]

INTRODUCTION

The management of patients with degenerative diseases of the spine requires reliable measures of symptoms, function, and outcome. Choice of conservative or surgical therapy in this cohort is complex and depends on many factors, including the history, quality, and severity of pain; functional limitations; and health-related quality of life (HRQOL) (1,2). An accurate and thorough evaluation of both the subjective and objective condition is helpful to select adequate treatment for the individual patient (3). In recent years, subjective patient-reported outcome measures (PROMs) in the form of questionnaires have been established as the gold standard for the clinical evaluation of spine patients (4). General pain measures, including the visual analog scale (VAS) or the numeric rating scale (NRS) for low back pain (LBP) or irradiating extremity (arm/leg) pain, are among the most commonly used PROMs. However, they are generic and not specific tools for spinal diseases (5). The Oswestry Disability Index (ODI), which is included in this review, is a prime example of an in-depth validated PROM and represents one of the most established instruments for a variety of different diseases of the lumbar spine (6,7). There is a broad variety of degenerative spine diseases, some of which are unspecific, whereas others (eg, lumbar spinal stenosis [LSS]) present with characteristic symptoms (eg, neurogenic claudication) that can be addressed by disease-specific tools (8). Cervical or thoracic degenerative spine disease may present with both radicular pain and/or myelopathy; outcome measures have to account for these different clinical manifestations (9).

Apart from subjective PROMs, objective measures of function are gaining increasing attention in spine research and have found their way into clinical practice (8,10). Measurements like the Timed Up-and-Go (TUG) test and the motorized treadmill test (MTT) assess a patient's objective functional impairment (OFI) and add a new dimension to the comprehensive patient evaluation (11). Because patients prefer objective functional tests rather than questionnaires and considering the continuous validation and standardization of objective outcome measures, we included the two most frequently applied tests of this relatively new field of outcome assessment in this review (8,12).

Degenerative disease of the spine encompasses a wide range of different pathologies and disease-specific symptoms. This leads to an even greater number of outcome measures that cannot all be included in the scope of this review. However, this article should provide the reader with a comprehensive summary of carefully selected instruments (Tables 1 and 2) (5,8,13).

PATIENT-REPORTED OUTCOME MEASURES

SPINAL STENOSIS MEASURE (SSM)

Description

Purpose. The SSM was developed in 1995 by Stucki et al as a short self-administered questionnaire to assess pain-related disability and health-related parameters in patients diagnosed with LSS (14). The SSM is also known as the Zurich Claudication Questionnaire, the Swiss Spinal Stenosis Questionnaire, or the Brigham Spinal Stenosis Questionnaire. The measure specifically addresses symptoms and functional deficits resulting from neurogenic claudication (14). It also includes an optional domain on patient satisfaction regarding the result of surgery. The SSM is one of the leading PROMs used by both spine surgeons and rheumatologists (5,15).

[1]Nicolai Maldaner, MD: University Hospital Zurich and University of Zurich, Zurich, Switzerland, and Cantonal Hospital St. Gallen, St. Gallen, Switzerland; [2]Martin Nikolaus Stienen, MD, FEBNS: University Hospital Zurich and University of Zurich, Zurich, Switzerland, and Stanford University Hospital and Clinics, Stanford, California.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Martin Nikolaus Stienen, MD, FEBNS, University Hospital Zurich, Department of Neurosurgery, Frauenklinikstrasse 10, Zurich 8091, Switzerland. E-mail: mnstienen@gmail.com.

Submitted for publication January 5, 2020; accepted in revised form April 2, 2020.

Content. The SSM contains three scales, with seven questions on symptom severity (SSM symptoms), five questions on physical function (SSM function), and six questions on satisfaction (SSM satisfaction). The SSM symptom scale encompasses a pain subdomain (three items) and a neuroischemic subdomain (four items).

Number of items. Eighteen items.

Response options/scale. Each item is measured on a three- to five-point ordinal scale, ranging from the best scenario to the worst scenario.

Recall period for items. The scale relates to symptoms experienced over the past month.

Cost to use. None.

How to obtain. The items are reported in the original publication (14). A copy is provided as an appendix to this article (English version).

Practical application

Method of administration. The SSM is self-administered by patients using paper and pen. Administration by telephone is also possible.

Scoring. The symptom severity score is the unweighted mean of the first seven questions; the resulting score indicates greater symptom severity. The first six items are scored from 1 to 5, whereas the seventh item has only three possible responses corresponding to scores of 1, 3, or 5. The symptom severity score can be further divided into a subscale for pain (questions one through three) and neuroischemic symptoms (questions four through seven). The physical function score is calculated as the unweighted mean of the following five questions, all scored from 1 to 4, with higher scores representing greater disability. The third scale relates to patient satisfaction after treatment, with higher scores indicating greater satisfaction. The unweighted mean is only valid if no more than one item is missing in the SSM function and SSM satisfaction scores and no more than two items are missing in the SSM symptom score. Some authors prefer to depict the SSM as the sum score without taking the unweighted mean.

Score interpretation. An unweighted mean is calculated for all three subscales. The SSM symptom severity score ranges from 1 (no symptoms) to 5 (maximal symptoms), which is further divided into the pain subscales (scores between 1 and 5) and neuroischemic symptom subscale (scores between 1 and 5). The SSM functions score ranges from 1 (no disability) to 4 (maximum disability), and the SSM satisfaction score ranges from 1 (very satisfied) to 4 (very dissatisfied). When the SSM is declared as a sum score, the SSM symptom score ranges from 7 (no symptoms) to 35 (maximal symptoms), the SSM functions score ranges from 5 (no disability) to 20 (maximum disability), and SSM satisfaction score ranges from 6 (very satisfied) to 24 (very dissatisfied).

Respondent time to complete. The SSM is relatively short and simple to fill out. It can be completed by the patient in less than 5 minutes.

Administrative burden. The SSM requires little administrative burden. The time to score (by hand) is short. No training is required.

Translations/adaptations. Translations are available in German and English. There are published adaptations in Spanish, Chinese, French, Polish, Norwegian, Iranian, Korean, Slovenian, and other languages, which showed satisfactory to excellent reliability and validity (15–23).

Psychometric information

Floor and ceiling effects. The authors are not aware of any studies reporting on floor or ceiling effects of the scales.

Reliability. The reliability in test-retest assessment consistently scores of more than 80% to 90% agreement (24–26).

Validity. The SSM has been validated against the self-paced walking test, the ODI, the Oxford Spinal Stenosis Score, the Short Form 36 (SF-36), the Health Utilities Index, the Center for Epidemiologic Studies Depression Scale, and other measures with high correlations (25–27). Several studies have supported the construct validity of the SSM and psychometric properties and physical function in the LSS population (25,27,28).

Responsiveness. Standardized response means (SRMs) are reported to be 1.07 for the SSM function scale and 0.96 for the SSM symptom scale in the original study with 130 patients (28). In a subsequent independent study among 91 patients with LBP, the SRM of SSM function (1.00) and SSM symptom (1.43) were confirmed, and the composite SSM average score even increased in responsiveness (1.50) (29). Because of its disease specificity, the SSM has been shown to be more responsive than the Sickness Impact Profile and the Roland-Morris Disability Questionnaire (RMDQ) in assessing patients with LSS (28).

Minimal clinically important differences. Stucki et al reported a difference in the physical function scale and symptom severity scores between the unsatisfied and somewhat satisfied patients of 0.52 (13%) and 0.48 (9.6%) (14).

Generalizability. The SSM is specific to symptoms and functional deficits characteristic of neurogenic claudication. For LSS, excellent generalizability has been proven by consistent results across multicultural studies (16–23).

Use in clinical trials. The SSM has been used in various recent studies and clinical trials (30–33).

Critical appraisal of overall value to the rheumatology community

Strengths. The SSM is a short disease-specific questionnaire for patients with LSS with three different subscales, namely, symptom severity, physical function, and satisfaction. The SSM proved to be the most precise in patients with LSS when compared with the ODI or the Oxford Claudication Score (25).

Caveats and cautions. A publication by Comer et al used Rasch analysis to evaluate the psychometric properties of the SSM (24). In their study, the SSM symptom scale, by measuring pain and neuroischemic symptoms as two separate constructs, failed to function as a unidimensional domain. Because of this, the SSM symptom scale was further subdivided into the two separate subscales. The SSM function proved valid for group comparison, although question 11 was not considered clinically meaningful (24).

Clinical usability. The SSM has established psychometric properties, and both administrative and respondent burdens are low. It is frequently used to assess and monitor outcome. As the current gold standard of outcome assessment in patients with LSS, it has been endorsed by the North American Spine Society and termed "best and most specific outcome measure for LSS" (34).

Research usability. The SSM is widely used in clinical research. The SSM symptom scale should be further subdivided into the SSM pain symptom and SSM neuroischemic symptom scales to adequately address these separate conditions.

CORE OUTCOME MEASURES INDEX BACK (COMI BACK)

Description

Purpose. Ever since a multinational group of experts proposed the COMI as a standardized outcome assessment in LBP research in 1998, this short multidimensional outcome measure has been thoroughly validated. It is widely used, especially to monitor outcomes in patients with various spinal disorders who are undergoing surgery or any form of intervention (35–37). Initially proposed as a set of six questions, a seventh item (overall quality of life) was added in 2005 by Mannion et al to construct what is known as the COMI Back (38). A validated and reliable version designed for cervical spine diseases, the Core Outcome Measures Index Neck, is also available (39,40).

Content. The COMI Back focuses on the patient's perspective (29). Following this premise, it covers the domains of pain intensity (LBP and leg/buttock pain), back-related function, symptom-specific well-being, general quality of life, and social/work disability. A further part covers patient satisfaction after surgical treatment.

Number of items. Seven items. (There are an additional four questions at follow-up that cover patient satisfaction and treatment results.)

Response options/scale. The first two questions regarding pain intensity are measured on a 0 to 10 NRS. The other five items are measured on a five-point Likert scale ranging from "best" to "worst."

Recall period for items. All items refer to the last week (except for disability, which refers to the last 4 weeks).

Cost to use. None.

How to obtain. A copy in different languages (including English, German, French, and Spanish) can be downloaded from the EUROSPINE Spine Tango registry website (https://www.eurospine.org/forms.htm).

Practical application

Method of administration. The COMI Back is self-administered by patients using paper and pen or online forms.

Scoring. The higher of the two pain scores (scores of 0-10) is taken as the pain intensity score. The other six items on the five-point Likert scales correspond with scores of 0, 2.5, 5.0, 7.5, and 10, respectively. The average of the two disability items form the disability score. The five domain scores for pain intensity, back-related function, symptom-specific well-being, general quality of life, and disability are then averaged to give a COMI Back score that ranges from 0 to 10.

Score interpretation. Scores range from 0 (best) to 10 (worst).

Respondent time to complete. The COMI Back is relatively short and simple to fill out. It can be completed by the patient in less than 5 minutes.

Administrative burden. The COMI Back requires little administration burden. The time to score (by hand) is short. No training is necessary.

Translations/adaptations. The COMI Back is available and cross-culturally validated in many languages, including English (38), German (38), Korean (41), Italian (42), French (43), Turkish (44), and Japanese (45) among others.

Psychometric information

Floor and ceiling effects. In one of the original articles, floor or ceiling effects in the range of 20% to 50% were observed for some items of the COMI Back before surgery (function and symptom-specific well-being) and after surgery (disability and function) (40).

Reliability. Several research groups examined the psychometric properties of the COMI Back in patients with various pathologies presenting with LBP. The test-retest reliability consistently scored highly, with an intraclass correlation coefficient (ICC) of 0.8 to 0.9 or more. High internal consistency for the core item index was shown for patients with chronic LBP (Cronbach's $\alpha = 0.90$ or more); however, less was shown in patients with acute osteoporotic fractures (Cronbach's $\alpha = 0.64$ or more) (38,40,46).

Validity. Good construct validity of the score was demonstrated by a moderate to high correlation with reference questionnaires, including the RMDQ, the SF-36, and the ODI ($r = 0.60$-0.79) (38,46). The item symptom-specific well-being, however, showed little correlation to other measures ($r = 0.25$-0.31), which might indicate that this item delivers unique information that may be of importance to the multidimensional nature of the overall index (38).

Responsiveness. Internal responsiveness corresponding with the SRM at the 12-month follow-up in a cohort of 91 patients undergoing surgery for LSS showed great responsiveness (SRM 1.44), which was similar to or better than the SSM average (1.50), the RMDQ (1.13), and the NRS pain (1.28) (29). In the same cohort, external responsiveness—meaning the strength of correlation between its change in scoring and the change in other outcome instruments—showed moderate correlation with the SSM ($r = 0.62$) and RMDQ ($r = 0.43$). In an area under the receiver operating characteristics (AUROC) curve analysis, good discriminative ability between good and bad outcomes was demonstrated (area under the curve of more than 0.83) (30).

Minimal clinically important differences. One of the original studies examining the COMI Back in 277 patients with LBP reported a minimum clinically important difference for improvement ($MCID_{imp}$) as a two- to three-point decrease, depending on the anchor used to indicate treatment success (38). In a group of 3056 patients undergoing spine surgery for a variety of indications, the groups' mean $MCID_{imp}$ at 12 months was –2.6 points, and the minimum clinically important difference for deterioration ($MCID_{det}$) was 1.2 points. AUROC curves of 0.88 for the $MCID_{imp}$ and 0.89 for the $MCID_{det}$ indicated good discriminative ability. The cutoffs for individual improvement and deterioration were –2.2 points or less (sensitivity 81%, specificity 83%) and 0.3 points or more (sensitivity 83%, specificity 88%), respectively (36).

Generalizability. Psychometric properties and sensitivity to change of the COMI Back are strong in patients with LBP from a variety of causes (38,40,46,47). Thus, it can be applied to a broad range of patients with LBP.

Use in clinical trials. The COMI Back has been used in multiple clinical trials (48–50).

Critical appraisal of overall value to the rheumatology community

Strengths. The COMI Back is a short but independently and repeatedly validated outcome measure for a variety of different pathologies presenting with LBP and leg pain. It is easily accessible and is associated with a low burden for both patients and physicians.

Caveats and cautions. As in other back-specific outcome measures (51,52), the COMI Back's values for the $MCID_{det}$ are lower than those for the $MCID_{imp}$, indicating that the COMI is less responsive to deterioration than to improvement (36).

Clinical usability. The COMI Back has shown great psychometric properties and is brief enough to be practical for routine clinical use and quality management. Based on these values, it has been incorporated into the European EUROSPINE registry as an outcome questionnaire of choice (53).

Research usability. The availability and ease of administration is meant to encourage clinicians and researchers to collaborate in registries and research projects on a national and international level (54). However, on an international level, outcomes measures like the ODI and RMDQ are still more established (5).

NECK DISABILITY INDEX (NDI)

Description

Purpose. Outcome research in cervical spine patients is historically less developed compared with lumbar spine patients (5,55). The NDI is one of the best established and most commonly used PROMs for chronic neck pain (56,57). Originally published in

1991, it is also known as the Vernon-Mior Disability Index or the Neck Pain Disability Index (58,59).

Content. The NDI measures pain and disability in patients with neck pain. The NDI covers the domains of pain intensity, personal care, lifting, work, headache, concentration, sleeping, driving, reading, and recreation.

Number of items. Ten items.

Response options/scale. Each of the 10 items are scored on a six-point Likert scale ranging from 0 (no pain/disability) to 5 points (maximal pain/disability).

Recall period for items. The questions refer to the current clinical condition and pain intensity. The exact time frame is not defined.

Cost to use. None.

How to obtain. Copies in English and other languages can be found in published sources or from the EUROSPINE Spine Tango registry website (https://www.eurospine.org/forms.htm) (56,58,59).

Practical application

Method of administration. The NDI is self-administered by patients using paper and pen.

Scoring. All 10 items are scored from 0 to 5. Numeric responses are summed up to a total score ranging from 0 (best) to 50 (worst). The developer and others recommend scoring the NDI out of 50 points. If three or more items are missing, the questionnaire is not valid. If two or fewer items are missing, the score should be normalized to 50 (58,59). Besides presenting raw test results, many groups have expressed the score as a percentage, ranging from 0% to 100% neck disability.

Score interpretation. No consensus exists regarding score interpretation. The original developers suggested the following: scores between 0 and 4 represent no disability, scores between 5 and 14 represent mild disability, scores between 15 and 24 represent moderate disability, scores between 25 and 34 represent severe disability, and scores greater than 35 represent complete disability (56).

Respondent time to complete. The NDI is relatively short and simple to fill out. It can be completed by the patient in 3 to 8 minutes (56).

Administrative burden. The NDI requires no administration burden and takes less than 3 minutes to score by hand. No training necessary.

Translations/adaptations. The NDI was originally developed in English but has been culturally adapted and translated into several languages including German (60,61), Spanish (62), Arabic (63), Chinese (64), Turkish (65), Japanese (66), Polish (67), Finnish (68), Greek (69), Portuguese, and other languages (70).

Psychometric information

Floor and ceiling effects. Patients with baseline score in the lower 10th and the upper 90th percentile are subject to significant floor or ceiling effects, respectively. This is why caution is required when using the NDI to monitor outcome in these high- or low-performing patients (56,71,72).

Reliability. A systematic review published in 2009 analyzed 41 studies that examined at least one aspect of the psychometric properties of the NDI. A high test-retest reliability was demonstrated in populations with both acute and chronic neck pain, with a reliability coefficient of more than 0.90 in most studies (56).

Validity. The NDI score correlates strongly (*r* of more than 0.70) with other neck disability measures, including the Neck Pain and Disability Scale, the Cervical Spine Outcome Questionnaire, the Disability Rating Index, and the VAS pain scale. Furthermore, high Cronbach's α scores of 0.70 to 0.96 show good internal consistency (56,57,60,70). The NDI is, however, only moderately correlated with both physical and mental aspects of general health as assessed by the SF-36 (56).

Responsiveness. The NDI has a good ability to detect changes over time. The SRM ranges from 0.60 to 0.95 in a systematic literature review (56).

Minimal clinically important differences. Reported MCIDs range from 5 of 50 to 19 of 50 points (56). Neck pain of musculoskeletal origin seems to show a lower MCID (change of more than 5 points) compared with nerve-related pain (7-13.4 points) (56,73,74). It is important to understand that each MCID is disease specific and, accordingly, shows slight variations, but on average a change of 5 to 7 points on the NDI (10%-14%) can be considered a clinically meaningful change in pain and disability.

Generalizability. Because of its strong psychometric properties and its cross-validation in several languages, the NDI can be used in a variety of different patient populations with neck pain.

Use in clinical trials. The NDI has been used in multiple clinical trials (75,76).

Critical appraisal of overall value to the rheumatology community

Strengths. The NDI is a reliable, valid, and responsive measure in various populations. This includes patients with both acute and chronic conditions as well as patients with neck pain associated with musculoskeletal dysfunction, whiplash-associated disorders, and cervical radiculopathy (56,60).

Caveats and cautions. The MCID differs across study populations and publications, which complicates the standardized interpretation of NDI results over time in clinical practice and research (56). The NDI does not include psychosocial and emotional aspects, although these are quite common in patients with chronic neck pain (56). We recommend scoring the NDI out of 50 points, as originally proposed.

Clinical usability. The NDI is one of the most common PROMs for neck pain and is regularly used by rheumatologists, spine surgeons, and physiotherapists (13,56,57). More studies reporting MCIDs for specific pathologies and patient populations are required (56).

Research usability. The good psychometric properties support using the NDI in research on cervical spine or neck diseases.

OSWESTRY DISABILITY INDEX

Description

We would like to refer to the article by Smeets et al (77) for a comprehensive description of the ODI. To avoid considerable redundancy, we provide only a brief overview and literature update on the ODI.

Purpose. The ODI assesses pain-related disability in patients with LBP associated with a wide range of causes/disorders. It is especially useful in patients with severe, persistent disability (6). Initially published in 1980 by John O'Brian (version 1.0), it has been modified several times (6,78,79). All versions of the ODI remain in use; however, not all have been subject to the same systematic validation as the original. The ODI version 2.0 developed by Fairbank et al is recommended for general use and has been adapted by various spine societies (6,7).

Content. The ODI contains one item on pain intensity and nine items on activities of daily living (ADLs) (personal care, lifting, walking, sitting, standing, sleeping, sex life, social life, and traveling).

Number of items. Ten items.

Cost to use. No permission is required, and the questionnaire is free.

How to obtain. The ODI can be found in the original article (7).

Psychometric information

The ODI has been used in multiple clinical trials (32,80,81).

Critical appraisal of overall value to the rheumatology community

Clinical usability. The ODI is one of the most commonly used clinical measures of pain-related disability in a variety of pathologies of the lower back. The ODI has established psychometric properties, a high reliability and validity, as well as low respondent and administrative burden (77).

Research usability. The ODI is one of the most frequently used outcome questionnaires for assessing and monitoring outcomes in both patients undergoing surgical or conservative care (5,82). It is frequently used as a comparator to validate other outcome measures.

ROLAND-MORRIS DISABILITY QUESTIONNAIRE

Description

Similar to the ODI, we would like to refer to the article by Smeets et al (77) for a more comprehensive description of the RMDQ.

Purpose. Originally designed in 1983, the RMDQ assesses physical disability due to LBP (83). It has been extensively used for acute, subacute, and chronic LBP as well as sciatica in various clinical settings (84). Since its first publication, several modifications have been proposed to improve suitability for specific patients (eg, for those with sciatica). However, changes resulted only in minor improvements and reduced comparability (85). The use of the original 24-item RMDQ is therefore recommended (7,35).

Content. The RMDQ assesses the execution of ADLs and common functions (housework, sleeping, dressing, mobility, appetite, etc). It does not exclusively contain items to rate physical disability but rates general impairment and health as well.

Number of items. Twenty-four items.

Cost to use. No permission is required, and the questionnaire is free.

How to obtain. Free download in different languages is available on the RMDQ website (www.rmdq.org/). A copy is attached in the original publication (7).

Psychometric information

The RMDQ has been used in multiple clinical trials (86–88).

Critical appraisal of overall value to the rheumatology community

Clinical usability. Although it is nonspecific to any spine pathology, the RMDQ is one of the most established and thoroughly evaluated measures for patients with LBP. It has psychometric properties ranging from acceptable to good. However, scores and changes must be interpreted with caution because some items are poor-fitting, and the overall score does not have interval-level properties.

Research usability. Psychometric qualities are acceptable for use in research. Rasch analysis and examination of score distribution is recommended as part of statistical analysis (89) (see Tables 1 and 2).

OBJECTIVE OUTCOME MEASURES

TIMED UP-AND-GO TEST

Description

Purpose. The TUG is the most frequently applied objective task-based functional outcome measure in patients with lumbar degenerative disk diseases (DDDs) (8). It includes several activities that are difficult to perform for patients with painful conditions or disability resulting from lumbar DDD, eg, standing up, walking fast, changing direction, and sitting down (8,90).

Content. The patient is asked to sit and lean back on an armchair, with arms resting on the armrests. On command, the patient should get up and walk as fast as possible to a line marked on the floor at 3 meters away from the chair. At this line, the participant should turn around, walk back to the chair, and sit down as quickly as possible. The examiner records the time (in seconds) between getting up and sitting down again.

Cost to use. None.

How to obtain. Measurement is possible with any commercial stopwatch. A smartphone application (TUG app) is available in multiple languages and conveniently measures and scores the TUG test based on normal population reference values. It can be downloaded free of cost in the Apple App Store (https://itunes.apple.com/de/app/tug-app/id1119087707?mt=8) and the Android Play Store (https://play.google.com/store/apps/details?id=ch.webgearing.tugapp).

Practical application

Method of administration. The TUG test does not require any special equipment except for a chair, a timer, and 3 meters of unrestricted walking space. It is currently recommended that an examiner takes the patient's time; however, patient self-measurements have been found to be reliable for similar objective functional tests (91). Walking aids (cane, walker, etc) are permitted if required.

Scoring. Results are traditionally expressed as raw test times (in seconds). Because raw TUG test times naturally vary, it is recommended to express TUG test results in a standardized manner for age and sex. Standardized OFI Z scores/T scores express a patient's deviation from the normal population mean and can be calculated conveniently using the TUG app (see How to obtain) (11,92–95).

Score interpretation. Expressing OFI as a standardized Z/T score prevents bias introduced by the high influence of the variables age and sex and can express disability on an interval scale (93,96,97). Alternatively, TUG test results can be stratified according to severity into no OFI (TUG test results are less than the upper limit of the normal population range [ULN]), mild OFI (the TUG test results are between the ULN and the 33rd percentile of the diseased population), moderate OFI (TUG test results are between the 33rd and 66th percentile of the diseased population), and severe OFI (TUG test results are greater than the 66th percentile of the diseased population) (11,95).

Administrative burden. The TUG requires very low administrative burden. Completion typically takes less than 1 minute. The administrative burden is lower than that of most PROMs, and the TUG test was preferred over a set of questionnaire-based PROM assessments by 60% to 70% of patients (12).

Translations/adaptations. No translation is needed for the TUG test. It can be demonstrated to patients if language barriers exist.

Psychometric information

Floor and ceiling effects. For discriminating subjects with healthy spines and patients with spine disease, the ULN was set as the 99th percentile of the normal population's TUG test results (11). In subsequent series that examined surgical candidates with lumbar DDD, a considerable proportion of candidates (approximately 60%) scored within the normal population range (no OFI). Thus, the standardized interpretation of TUG test values is likely to have a floor effect (11,95). There is no upper time limit for the TUG test, and the calculation of standardized OFI Z/T scores allows for the accurate determination of extreme disability without a ceiling effect.

Reliability. The TUG showed an excellent intrarater (ICC = 0.97) and interrater reliability (ICC = 0.99) with an SEM of 0.21 to 0.23 sec (11).

Validity. Adequate convergent validity with PROMs, including VAS back (r = 0.25) and VAS leg pain (r = 0.29), RMDI (r = 0.38), and ODI (r = 0.34) as well as the Short Form 12 (SF-12) physical component summary (r = –0.32) and EuroQol Five Dimension Questionnaire (EQ-5D) (r = –0.28) could be demonstrated (11).

Responsiveness. Several studies demonstrated that the TUG test is sensitive to the postoperative change in function of a patient with DDD (92,94,98). The SRM is 0.51 in a sample of 123 patients with lumbar DDD who were examined before and 6 weeks after lumbar spine surgery (data not published).

Minimum clinically important differences. In patients with lumbar DDD, a change in raw TUG test time of at least 3.4 s is considered a clinically meaningful change in function (94).

Generalizability. The TUG test has been used in a broad variety of pathologies of the lumbar spine (8). Only a few studies so far have focused on the psychometric properties of the TUG for a specific disease (eg, LSS) (99).

Use in clinical trials. The TUG has been used in multiple clinical trials (88,90,100,101).

Critical appraisal of overall value to the rheumatology community

Strengths. The TUG test is safe, fast, easy to conduct, and well appreciated by patients (12). It is the most thoroughly validated objective outcome measure to assess OFI in patients with degenerative disease of the lumbar spine (8). Its test result is (relatively) independent from the confounding influence of the variables age and sex (11), body mass index (101), and mental health status (102), but it takes disability resulting from lower extremity motor deficits into account (103).

Caveats and cautions. The test cannot be applied to patients who are unable to ambulate (despite walking aid) because of severe pain or neurological deficits. There may be a significant influence of comorbidities (eg, hip or knee osteoarthritis, congestive heart failure, Parkinson's disease, etc), which renders interpretation of test results in these patients difficult.

Clinical usability. Quick to conduct and interpret without the need for special equipment, the TUG test can easily be integrated into clinical routine (administration by nursing staff, advanced practice providers/physician assistants, or physical therapists [PTs]). It should be noted that the TUG test correlated moderately with a broad range of different pain, disability, and HRQOL measures, indicating that the TUG test can provide a broad impression about the patient's general functional status. However, for an in-depth analysis of specific aspects of a patient's condition, the assessment may need to be supplemented with one or several PROMs (11).

Research usability. The strong psychometric properties of reliability and validity paired with the low administrative burden and high patient acceptance demonstrate that the TUG is an accurate and quick tool for the measurement of OFI in clinical spine research.

MOTORIZED TREADMILL TEST (MTT)

Description

Purpose. In a recent systematic review, the MTT was the second-most frequently applied measure to determine OFI in patients with lumbar DDD (8). Because of the typically long treadmill protocols that challenge the walking capacity of participants, it has been primarily studied in patients with LSS and neurogenic claudication. The MTT has been used as an objective outcome measure in several randomized controlled trials and observational studies (8).

Content. The participant is instructed to walk on a calibrated treadmill, usually at a predefined protocol starting on a level surface. Different protocols in terms of speed, time, or incline exist without an accepted gold standard (8). Ambulation time and distance, walking speed, and time of onset or significant increase in symptoms are monitored as test results. Several studies proposed a protocol that starts with 10 minutes at 2 mph, increases to 2.5 mph for the next 5 minutes, and then increases to 3 mph for an additional 5 minutes (total of 20 minutes) (104,105). Some protocols remain at a constant speed between 2 and 2.5 mph for the complete duration of the test (between 15 and 30 min-

utes) (106–110), whereas others instruct participants to walk at a maximum individually selected speed for up to 15 to 30 minutes (111–114). Further individualized protocols exist (8).

Cost to use. The MTT requires the cost of special equipment (motorized treadmill), and PTs are needed to safely administer the test.

How to obtain. MTT protocols can be adopted using published literature. For different protocols, see the references above.

Practical application

Method of administration. The MTT requires a programmable motorized treadmill and trained personal to supervise protocol adherence and patient safety. In most programs, participants are allowed to hold one handrail for balance purposes. Using both handrails is restricted because it can improve walking capacity and speed in LSS by allowing the patient to bending forward (8,112,115). In the case that a patient does not tolerate the standard speed, it is reduced by the supervisor or the test is ended. Some studies use safety end points (eg, 85% of predicted maximal heart rate [220 – age]) as an additional reason for premature test termination (116). Some authors state that the additional information gained after 15 minutes of test time is negligible (117,118).

Scoring. Test results derive from 1) the time to first symptoms (TTFS) for general pain and/or paresthesia, 2) total ambulation time (in minutes or seconds) and distance (in meters), and 3) maximum walking speed (in m/s).

Score interpretation. All published studies so far have reported raw test results. No studies have expressed results in standardized fashion using Z- or T scores. Moreover, different walking protocols impede result comparisons across different studies (8).

Administrative burden. Test time is 15 to 30 minutes. In addition, the patient needs to be instructed in written or verbal form, and some studies use a heart rate monitoring system. According to the specific walking protocol, results in different stages of the test can be recorded.

Translations/adaptations. No translation is needed for the MTT.

Psychometric information

Floor and ceiling effects. The authors are unaware of studies reporting on the MTT's floor or ceiling effects. However, a relatively poor discriminative capacity between subjects with healthy spines and patients with spine disease with low disability burden can be assumed (similar to the TUG test). A study by Tomkins et al showed that the MTT likely underestimates a patient's true walking capacity when compared with a different walking test (the self-paced walking test [SPWT]) (119).

Reliability. The MTT showed a high to excellent intrarater reliability, both for TTFS (ICC = 0.90-0.98) and total ambulation time (ICC = 0.89-0.96) at 1.2 mph fixed or individually selected walking speed (120). Another group reported slightly lower intrarater reliability (ICC = 0.83) with a different protocol that simulated a gradual increase in walking speed (121).

Validity. Convergent validity of the MTT could be demonstrated when compared with the SPWT (r = 0.88) (119) and self-reported symptoms of neurogenic claudication (r = 0.88) (104) as well as ODI scores (r = –0.51) and self-reported walking distance (r = 0.62) (106).

Responsiveness. Several studies have documented the MTT's good responsiveness to change in conservatively and surgically treated patients with lumbar spine disease; however, no SRMs are reported and between-study comparisons are complicated by different MTT protocols (117,122,123). Moreover, in a report of 32 patients undergoing surgery or conservative treatment for LSS, internal responsiveness was low and was significantly worse compared with the SPWT (effect size [ES] in time 0.17 min; ES in distance 0.09 m; ES in speed 0.11 km/h) (112).

Minimum clinically important differences. The authors are unaware of any studies reporting on the MTT's MCIDs.

Generalizability. Most studies applied the MTT to patients with LSS because of its ability to trigger symptoms and functional deficits characteristic of neurogenic claudication. Within the LSS population, good psychometric properties have been shown. However, the great variety of test protocols complicates generalizability of the results.

Use in clinical trials. The MTT has been used in multiple clinical trials (109,110,116).

Critical appraisal of overall value to the rheumatology community

Strengths. The MTT helps objectify functional impairment in LSS over time as well as before and after conservative or surgical treatment. It is validated and used in numerous studies.

Caveats and cautions. The MTT cannot be applied to patients who are unable to ambulate. Other comorbidities, especially hip/knee osteoarthritis and severe pulmonary and cardiac diseases, may represent significant confounders by limiting walking

capacity or by making this exercise-based test medically inadvisable. A certain risk of frightening or even injuring patients who are elderly needs to be considered and might lead to test dropouts (124).

Clinical usability. As a measure of walking capacity, the MTT can provide a valuable impression of a patient's functional status, thereby supplementing PROM assessments, especially in patients with LSS. A drawback for the clinical use of the MTT is that it is comparably resource-intensive, requiring special equipment and trained personnel.

Research usability. There is sufficient evidence of the reliability and validity of the test. The MTT might increase our insight into the interaction of objective physiological walking capacity and subjective PROMs. However, the lack of a widely accepted standardized protocol and insufficient data on responsiveness and MCIDs hamper research usability and comparison between studies.

CONCLUSIONS

PROMs remain the current gold standard in evaluating patients with degenerative diseases of the spine. Ideally, they should become an integral part of any institution's patient evaluation, both for clinical practice and research (125,126). PROMs help to evaluate, monitor, and compare treatment results over time and across populations.

The considerable variability of spinal pathologies and their clinical presentation led to the development of various disease-specific outcome measures, some of which are reviewed in this article (Tables 1 and 2). To remain concise and to not exceed the scope of this review, the authors acknowledge that there are several other important outcome measures that are not discussed. These include, for example, the Scoliosis Research Society Questionnaire 22, 23, and 30, assessing patients with spinal deformities; the (modified) Japanese Orthopaedic Association Myelopathy Scale, assessing patients with cervical spondylotic myelopathy; or the Copenhagen Neck Functional Disability Scale, assessing the level of functional disabilities in patients with neck pain. Other LBP-specific measures are the Low Back Pain Bothersome Scale, the Low Back Pain Impact Questionnaire, and the National Institutes of Health Low Back Pain Minimal Dataset. In addition, generic outcome measures (that are not disease-specific instruments for spinal pathologies), such as the Patient-Reported Outcome Measurement Information System measures, or HRQOL measures, such as the SF-12/SF-36 or the EQ-5D, have been extensively reviewed in the literature elsewhere (126,127). Other objective measures of function that exceed the scope of this paper but are systematically reviewed elsewhere include the SPWT, the 6-minute walking test (6WT), the five-repetition sit-to-stand test, the shuttle walking test, and the Short Physical Performance Battery besides step counters (8,10,128,129).

It is clear that certain spinal diseases may require specific assessment tools. However, to enable comparison of pain and disability between individuals, disease populations, or studies, in the vast majority of patients, the use of a single PROM would suffice. Because of its good psychometric properties, popularity, and ubiquitous use, we recommend use of the ODI as the PROM of first choice in patients with degenerative diseases of the lumbar spine. For the same reasons, we recommend using the NDI in patients with degenerate cervical spine diseases presenting with acute or chronic neck pain.

Regarding the evaluation of pain/disability over time, a baseline assessment, followed by 3-month and 12-month follow-up assessments after the initiation of treatment appears reasonable. Recent research indicates that most functional recovery after surgery can be expected within 8 to 12 weeks (10).

The broader availability of modern technologies such as smartphones or wearable devices with global positioning systems (GPS), combined with patient demand for a more personalized and transparent care, drive the trend to construct detailed, accurate, and intelligible medical profiles based on objective activity data (10,130). Tests such as the TUG test, the MTT, or the SPWT are already established and validated objective outcome measures used in spine care and research and can complement PROMs and add a further dimension to an in-depth patient evaluation. Free smartphone applications that accurately measure OFI by determining walking capacity based on GPS coordinates (eg, the 6WT app) are currently validated in clinical trials and offer great potential for further advances in this field (129).

AUTHOR CONTRIBUTIONS

All authors drafted the article, revised it critically for important intellectual content, and approved the final version to be published.

REFERENCES

1. Luijsterburg PA, Verhagen AP, Ostelo RW, van Os TA, Peul WC, Koes BW. Effectiveness of conservative treatments for the lumbosacral radicular syndrome: a systematic review. Eur Spine J 2007;16:881–99.
2. Van Hooff ML, van Loon J, van Limbeek J, de Kleuver M. The Nijmegen decision tool for chronic low back pain: development of a clinical decision tool for secondary or tertiary spine care specialists. PLoS One 2014;9:e104226.
3. Jespersen AB, Gustafsson ME. Correlation between the Oswestry Disability Index and objective measurements of walking capacity and performance in patients with lumbar spinal stenosis: a systematic literature review. Eur Spine J 2018;27:1604–13.
4. Mannion AF, Balagué F, Pellisé F, Cedraschi C. Pain measurement in patients with low back pain. Nat Clin Pract Rheumatol 2007;3:610–8.
5. Guzman JZ, Cutler HS, Connolly J, Skovrlj B, Mroz TE, Riew KD, et al. Patient-reported outcome instruments in spine surgery. Spine (Phila Pa 1976) 2016;41:429–37.
6. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine (Phila Pa 1976) 2000;25:2940–52.

7. Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire [published erratum appears in Spine 2001;26:847]. Spine (Phila Pa 1976) 2000;25:3115–24.
8. Stienen MN, Ho AL, Staartjes VE, Maldaner N, Veeravagu A, Desai A, et al. Objective measures of functional impairment for degenerative diseases of the lumbar spine: a systematic review of the literature. Spine J 2019;19:1276–93.
9. Misailidou V, Malliou P, Beneka A, Karagiannidis A, Godolias G. Assessment of patients with neck pain: a review of definitions, selection criteria, and measurement tools. J Chiropr Med 2010;9:49–59.
10. Stienen MN, Rezaii PG, Ho AL, Veeravagu A, Zygourakis CC, Tomkins-Lane C, et al. Objective activity tracking in spine surgery: a prospective feasibility study with a low-cost consumer grade wearable accelerometer. Sci Rep 2020;10:4939.
11. Gautschi OP, Smoll NR, Corniola MV, Joswig H, Chau I, Hildebrandt G, et al. Validity and reliability of a measurement of objective functional impairment in lumbar degenerative disc disease: the Timed Up and Go (TUG) test. Neurosurgery 2016;79:270–8.
12. Joswig H, Stienen MN, Smoll NR, Corniola MV, Chau I, Schaller K, et al. Patients' preference of the Timed Up and Go Test or patient-reported outcome measures before and after surgery for lumbar degenerative disk disease. World Neurosurg 2017;99:26–30.
13. Ueda H, Cutler HS, Guzman JZ, Cho SK. Current trends in the use of patient-reported outcome instruments in degenerative cervical spine surgery. Global Spine J 2016;6:242–7.
14. Stucki G, Daltroy L, Liang MH, Lipson SJ, Fossel AH, Katz JN. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. Spine (Phila Pa 1976) 1996;21:796–803.
15. North American Spine Society. Evidence-based clinical guidelines for multidisciplinary spine care: diagnosis and treatment of degenerative lumbar spinal stenosis. URL: https://www.spine.org/Portals/0/assets/downloads/ResearchClinicalCare/Guidelines/LumbarStenosis.pdf.
16. Fokter SK, Yerby SA. Patient-based outcomes for the operative treatment of degenerative lumbar spinal stenosis. Eur Spine J 2006;15:1661–9.
17. Hara N, Matsudaira K, Masuda K, Tohnosu J, Takeshita K, Kobayashi A, et al. Psychometric assessment of the Japanese version of the Zurich Claudication Questionnaire (ZCQ): reliability and validity. PLoS One 2016;11:e0160183.
18. Hidalgo Ovejero AM, Menéndez García M, Bermejo Fraile B, García Mata S, Forcén Alonso T, Mateo Sebastián P. Cross-cultural adaptation of the Zurich Claudication Questionnaire: validation study of the Spanish version. An Sist Sanit Navar 2015;38:41–52.
19. Kim HJ, Lee YK, Kim DO, Chang BS, Lee CK, Yeom JS. Validation and cross-cultural adaptation of the Korean version of the Zurich Claudication Questionnaire in patients with lumbar spinal stenosis. Spine (Phila Pa 1976) 2018;43:E105–10.
20. Marchand AA, Tétreau C, O'Shaughnessy J, Descarreaux M. French-Canadian adaptation and validation of the Swiss Spinal Stenosis Questionnaire for patients with lumbar spinal stenosis. Spine (Phila Pa 1976) 2019;44:E487–93.
21. Thornes E, Grotle M. Cross-cultural adaptation of the Norwegian version of the spinal stenosis measure. Eur Spine J 2008;17:456–62.
22. Tomaszewski KA, Kłosiński M, Henry BM, Skinningsrud B, Kucharska E, Dudkiewicz Z, et al. Large prospective validation and cultural adaptation of the Polish version of the Swiss Spinal Stenosis Questionnaire for patients with lumbar spinal stenosis. Ann Agric Environ Med 2017;24:676–82.
23. Yi H, Wei X, Zhang W, Chen Z, Wang X, Ji X, et al. Reliability and validity of simplified Chinese version of Swiss Spinal Stenosis Questionnaire for patients with degenerative lumbar spinal stenosis. Spine (Phila Pa 1976) 2014;39:820–5.
24. Comer CM, Conaghan PG, Tennant A. Internal construct validity of the Swiss Spinal Stenosis Questionnaire: Rasch analysis of a disease-specific outcome measure for lumbar spinal stenosis. Spine (Phila Pa 1976) 2011;36:1969–76.
25. Pratt RK, Fairbank JC, Virr A. The reliability of the Shuttle Walking Test, the Swiss Spinal Stenosis Questionnaire, the Oxford Spinal Stenosis Score, and the Oswestry Disability Index in the assessment of patients with lumbar spinal stenosis. Spine (Phila Pa 1976) 2002;27:84–91.
26. Tomkins-Lane CC, Battié MC. Validity and reproducibility of self-report measures of walking capacity in lumbar spinal stenosis. Spine (Phila Pa 1976) 2010;35:2097–102.
27. Tomkins CC, Battié MC, Hu R. Construct validity of the physical function scale of the Swiss Spinal Stenosis Questionnaire for the measurement of walking capacity. Spine (Phila Pa 1976) 2007;32:1896–901.
28. Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. J Clin Epidemiol 1995;48:1369–78.
29. Mannion AF, Fekete TF, Wertli MM, Mattle M, Nauer S, Kleinstück FS, et al. Could less be more when assessing patient-rated outcome in spinal stenosis? [comparative study]. Spine (Phila Pa 1976) 2015;40:710–8.
30. Austevoll IM, Hermansen E, Fagerland M, Rekeland F, Solberg T, Storheim K, et al. Decompression alone versus decompression with instrumental fusion the NORDSTEN degenerative spondylolisthesis trial (NORDSTEN-DS); study protocol for a randomized controlled trial. BMC Musculoskelet Disord 2019;20:7.
31. Ghogawala Z, Dziura J, Butler WE, Dai F, Terrin N, Magge SN, et al. Laminectomy plus fusion versus laminectomy alone for lumbar spondylolisthesis. N Engl J Med 2016;374:1424–34.
32. Försth P, Ólafsson G, Carlsson T, Frost A, Borgström F, Fritzell P, et al. A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. N Engl J Med 2016;374:1413–23.
33. Zucherman JF, Hsu KY, Hartjen CA, Mehalic TF, Implicito DA, Martin MJ, et al. A multicenter, prospective, randomized trial evaluating the X STOP interspinous process decompression system for the treatment of neurogenic intermittent claudication: two-year follow-up results. Spine (Phila Pa 1976) 2005;30:1351–8.
34. Watters WC, Baisden J, Gilbert TJ, Kreiner S, Resnick DK, Bono CM, et al. Degenerative lumbar spinal stenosis: an evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis. Spine J 2008;8:305–10.
35. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, et al. Outcome measures for low back pain research: a proposal for standardized use. Spine (Phila Pa 1976) 1998;23:2003–13.
36. Mannion AF, Porchet F, Kleinstück FS, Lattig F, Jeszenszky D, Bartanusz V, et al. The quality of spine surgery from the patient's perspective. Part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. Eur Spine J 2009;18 Suppl 3:374–9.
37. Mannion AF, Porchet F, Kleinstück FS, Lattig F, Jeszenszky D, Bartanusz V, et al. The quality of spine surgery from the patient's perspective. Part 1. The Core Outcome Measures Index in clinical practice. Eur Spine J 2009;18 Suppl 3:367–73.
38. Mannion AF, Elfering A, Staerkle R, Junge A, Grob D, Semmer NK, et al. Outcome assessment in low back pain: how low can you go? [comparative study]. Eur Spine J 2005;14:1014–26.
39. Fankhauser CD, Mutter U, Aghayev E, Mannion AF. Validity and responsiveness of the Core Outcome Measures Index (COMI) for the neck. Eur Spine J 2012;21:101–14.

40. White P, Lewith G, Prescott P. The core outcomes for neck pain: validation of a new outcome measure. Spine (Phila Pa 1976) 2004;29:1923–30.
41. Kim HJ, Yeom JS, Nam Y, Lee NK, Heo YW, Lee SY, et al. Validation and cross-cultural adaptation of the Korean version of the Core Outcome Measures Index in patients with degenerative lumbar disease. Eur Spine J 2018;27:2804–13.
42. Mannion AF, Boneschi M, Teli M, Luca A, Zaina F, Negrini S, et al. Reliability and validity of the cross-culturally adapted Italian version of the Core Outcome Measures Index. Eur Spine J 2012;21 Suppl 6:S737–49.
43. Genevay S, Cedraschi C, Marty M, Rozenberg S, de Goumoëns P, Faundez A, et al. Reliability and validity of the cross-culturally adapted French version of the Core Outcome Measures Index (COMI) in patients with low back pain. Eur Spine J 2012;21130–7.
44. Çetin E, Çelik EC, Acaroğlu E, Berk H. Reliability and validity of the cross-culturally adapted Turkish version of the Core Outcome Measures Index for low back pain. Eur Spine J 2018;27:93–100.
45. Matsudaira K, Oka H, Oshima Y, Chikuda H, Taniguchi Y, Matsubayashi Y, et al. Development of the Japanese Core Outcome Measures Index (COMI): cross-cultural adaptation and psychometric validation. BMC Musculoskelet Disord 2018;19:71.
46. Ferrer M, Pellisé F, Escudero O, Alvarez L, Pont A, Alonso J, et al. Validation of a minimum outcome core set in the evaluation of patients with back pain. Spine (Phila Pa 1976) 2006;31:1372–9.
47. Mannion AF, Elfering A, Staerkle R, Junge A, Grob D, Dvorak J, et al. Predictors of multidimensional outcome after spinal surgery. Eur Spine J 2007;16:777–86.
48. Fekete TF, Haschtmann D, Kleinstück FS, Porchet F, Jeszenszky D, Mannion AF. What level of pain are patients happy to live with after surgery for lumbar degenerative disorders? [comparative study]. Spine J 2016;16 Suppl:S12–8.
49. Mannion AF, Impellizzeri FM, Leunig M, Jeszenszy D, Becker HJ, Haschtmann D, et al. EUROSPINE 2017 full paper award: time to remove our rose-tinted spectacles: a candid appraisal of the relative success of surgery in over 4500 patients with degenerative disorders of the lumbar spine, hip or knee. Eur Spine J 2018;27:778–88.
50. Mannion AF, Denzler R, Dvorak J, Müntener M, Grob D. A randomised controlled trial of post-operative rehabilitation after surgical decompression of the lumbar spine. Eur Spine J 2007;16:1101–17.
51. Campbell H, Rivero-Arias O, Johnston K, Gray A, Fairbank J, Frost H, et al. Responsiveness of objective, disease-specific, and generic outcome measures in patients with chronic low back pain: an assessment for improving, stable, and deteriorating patients. Spine (Phila Pa 1976) 2006;31:815–22.
52. Hägg O, Fritzell P, Nordwall A. Swedish Lumbar Spine Study Group. The clinical importance of changes in outcome scores after treatment for chronic low back pain. Eur Spine J 2003;12:12–20.
53. Melloh M, Staub L, Aghayev E, Zweig T, Barz T, Theis JC, et al. The international spine registry SPINE TANGO: status quo and first results. Eur Spine J 2008;17:1201–9.
54. Balagué F, Mannion AF, Pellisé F, Cedraschi C. Clinical update: low back pain. Lancet 2007;369:726–8.
55. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF. Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. Spine (Phila Pa 1976) 2002;27:515–22.
56. MacDermid JC, Walton DM, Avery S, Blanchard A, Etruw E, McAlpine C, et al. Measurement properties of the neck disability index: a systematic review. J Orthop Sports Phys Ther 2009;39:400–17.
57. Macdermid JC, Walton DM, Côté P, Santaguida PL, Gross A, Carlesso L, et al. Use of outcome measures in managing neck pain: an international multidisciplinary survey. Open Orthop J 2013;7:506–20.
58. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther 1991;14:409–15.
59. Vernon H. The Neck Disability Index: state-of-the-art, 1991-2008. J Manipulative Physiol Ther 2008;31:491–502.
60. Cramer H, Lauche R, Langhorst J, Dobos GJ, Michalsen A. Validation of the German version of the Neck Disability Index (NDI). BMC Musculoskelet Disord 2014;15:91.
61. Swanenburg J, Humphreys K, Langenfeld A, Brunner F, Wirth B. Validity and reliability of a German version of the Neck Disability Index (NDI-G). Man Ther 2014;19:52–8.
62. Andrade Ortega JA, Delgado Martínez AD, Almécija Ruiz R. Validation of the Spanish version of the Neck Disability Index. Spine (Phila Pa 1976) 2010;35:E114–8.
63. Shaheen AA, Omar MT, Vernon H. Cross-cultural adaptation, reliability, and validity of the Arabic version of Neck Disability Index in patients with neck pain. Spine (Phila Pa 1976) 2013;38:E609–15.
64. Wu S, Ma C, Mai M, Li G. Translation and validation study of Chinese versions of the Neck Disability Index and the Neck Pain and Disability Scale. Spine (Phila Pa 1976) 2010;35:1575–9.
65. Telci EA, Karaduman A, Yakut Y, Aras B, Simsek IE, Yagli N. The cultural adaptation, reliability, and validity of Neck Disability Index in patients with neck pain: a Turkish version study. Spine (Phila Pa 1976) 2009;34:1732–5.
66. Nakamaru K, Vernon H, Aizawa J, Koyama T, Nitta O. Crosscultural adaptation, reliability, and validity of the Japanese version of the Neck Disability Index. Spine (Phila Pa 1976) 2012;37:E1343–7.
67. Misterska E, Jankowski R, Glowacki M. Cross-cultural adaptation of the Neck Disability Index and Copenhagen Neck Functional Disability Scale for patients with neck pain due to degenerative and discopathic disorders: psychometric properties of the Polish versions. BMC Musculoskelet Disord 2011;12:84.
68. Salo P, Ylinen J, Kautiainen H, Arkela-Kautiainen M, Häkkinen A. Reliability and validity of the Finnish version of the Neck Disability Index and the modified Neck Pain and Disability Scale. Spine (Phila Pa 1976) 2010;35:552–6.
69. Trouli MN, Vernon HT, Kakavelakis KN, Antonopoulou MD, Paganas AN, Lionis CD. Translation of the Neck Disability Index and validation of the Greek version in a sample of neck pain patients. BMC Musculoskelet Disord 2008;9:106.
70. Cook C, Richardson JK, Braga L, Menezes A, Soler X, Kume P, et al. Cross-cultural adaptation and validation of the Brazilian Portuguese version of the Neck Disability Index and Neck Pain and Disability Scale. Spine (Phila Pa 1976) 2006;31:1621–7.
71. Hains F, Waalen J, Mior S. Psychometric properties of the Neck Disability Index. J Manipulative Physiol Ther 1998;21:75–80.
72. Riddle DL, Stratford PW. Use of generic versus region-specific functional status measures on patients with cervical spine disorders. Phys Ther 1998;78:951–63.
73. Cleland JA, Childs JD, Whitman JM. Psychometric properties of the Neck Disability Index and Numeric Pain Rating Scale in patients with mechanical neck pain. Arch Phys Med Rehabil 2008;89:69–74.
74. Cleland JA, Fritz JM, Whitman JM, Palmer JA. The reliability and construct validity of the Neck Disability Index and patient specific functional scale in patients with cervical radiculopathy. Spine (Phila Pa 1976) 2006;31:598–602.
75. Vleggeert-Lankamp CL, Janssen TM, van Zwet E, Goedmakers CM, Bosscher L, Peul W, et al. The NECK trial: effectiveness of anterior cervical discectomy with or without interbody fusion and

arthroplasty in the treatment of cervical disc herniation: a double-blinded randomized controlled trial. Spine J 2019;19:965–75.

76. Badhiwala JH, Witiw CD, Nassiri F, Akbar MA, Mansouri A, Wilson JR, et al. Efficacy and safety of surgery for mild degenerative cervical myelopathy: results of the AOSpine North America and international prospective multicenter studies. Neurosurgery 2019;84:890–7.
77. Smeets R, Köke A, Lin CW, Ferreira M, Demoulin C. Measures of function in low back pain/disorders: Low Back Pain Rating Scale (LBPRS), Oswestry Disability Index (ODI), Progressive Isoinertial Lifting Evaluation (PILE), Quebec Back Pain Disability Scale (QBPDS), and Roland-Morris Disability Questionnaire. Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S158–73.
78. Fairbank JC. The use of revised Oswestry Disability Questionnaire. Spine (Phila Pa 1976) 2000;25:2846–7.
79. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry Low Back Pain Disability Questionnaire. Physiotherapy 1980;66:271–3.
80. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019;2:1.
81. Fritz JM, Magel JS, McFadden M, Asche C, Thackeray A, Meier W, et al. Early physical therapy vs usual care in patients with recent-onset low back pain: a randomized clinical trial. JAMA 2015;314:1459–67.
82. Vavken P, Ganal-Antonio AK, Quidde J, Shen FH, Chapman JR, Samartzis D. Fundamentals of clinical outcomes assessment for spinal disorders: clinical outcome instruments and applications. Global Spine J 2015;5:329–38.
83. Roland M, Morris R. A study of the natural history of back pain. Part I. Development of a reliable and sensitive measure of disability in low-back pain. Spine (Phila Pa 1976) 1983;8:141–4.
84. Grotle M, Brox JI, Vøllestad NK. Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. Spine (Phila Pa 1976) 2004;29:E492–501.
85. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. Spine (Phila Pa 1976) 1995;20:1899–908.
86. Friedly JL, Comstock BA, Turner JA, Heagerty PJ, Deyo RA, Sullivan SD, et al. A randomized trial of epidural glucocorticoid injections for spinal stenosis. N Engl J Med 2014;371:11–21.
87. Krebs EE, Gravely A, Nugent S, Jensen AC, DeRonne B, Goldsmith ES, et al. Effect of opioid vs nonopioid medications on pain-related function in patients with chronic back pain or hip or knee osteoarthritis pain. JAMA 2018;319:872–82.
88. Stienen MN, Joswig H, Chau I, Neidert MC, Bellut D, Wälchli T, et al. Efficacy of intraoperative epidural triamcinolone application in lumbar microdiscectomy: a matched-control study. J Neurosurg Spine 2018;28:291–9.
89. Garratt AM, United Kingdom Back Pain Exercise and Manipulation Trial. Rasch analysis of the Roland Disability Questionnaire. Spine (Phila Pa 1976) 2003;28:79–84.
90. Kondo R, Yamato Y, Nagafusa T, Mizushima T, Hasegawa T, Kobayashi S, et al. Effect of corrective long spinal fusion to the ilium on physical function in patients with adult spinal deformity. Eur Spine J 2017;26:2138–45.
91. Staartjes VE, Beusekamp F, Schröder ML. Can objective functional impairment in lumbar degenerative disease be reliably assessed at home using the five-repetition sit-to-stand test? A prospective study. Eur Spine J 2019;28:665–73.
92. Gautschi OP, Joswig H, Corniola MV, Smoll NR, Schaller K, Hildebrandt G, et al. Pre- and postoperative correlation of patient-reported outcome measures with standardized Timed Up and Go (TUG) test results in lumbar degenerative disc disease. Acta Neurochir (Wien) 2016;158:1875–81.
93. Gautschi OP, Smoll NR, Joswig H, Corniola MV, Schaller K, Hildebrandt G, et al. Influence of age on pain intensity, functional impairment and health-related quality of life before and after surgery for lumbar degenerative disc disease. Clin Neurol Neurosurg 2016;150:33–9.
94. Gautschi OP, Stienen MN, Corniola MV, Joswig H, Schaller K, Hildebrandt G, et al. Assessment of the minimum clinically important difference in the Timed Up and Go Test after surgery for lumbar degenerative disc disease. Neurosurgery 2017;80:380–5.
95. Stienen MN, Smoll NR, Joswig H, Corniola MV, Schaller K, Hildebrandt G, et al. Validation of the baseline severity stratification of objective functional impairment in lumbar degenerative disc disease. J Neurosurg Spine 2017;26:598–604.
96. Gautschi OP, Corniola MV, Smoll NR, Joswig H, Schaller K, Hildebrandt G, et al. Sex differences in subjective and objective measures of pain, functional impairment, and health-related quality of life in patients with lumbar degenerative disc disease. Pain 2016;157:1065–71.
97. Gautschi OP, Smoll NR, Corniola MV, Joswig H, Schaller K, Hildebrandt G, et al. Sex differences in lumbar degenerative disc disease. Clin Neurol Neurosurg 2016;145:52–7.
98. Gautschi OP, Corniola MV, Joswig H, Smoll NR, Chau I, Jucker D, et al. The Timed Up and Go Test for lumbar degenerative disc disease. J Clin Neurosci 2015;22:1943–8.
99. Stienen MN, Maldaner N, Joswig H, Corniola MV, Bellut D, Prömmel P, et al. Objective functional assessment using the "Timed Up and Go" test in patients with lumbar spinal stenosis. Neurosurg Focus 2019;46:E4.
100. Van Meirhaeghe J, Bastian L, Boonen S, Ranstam J, Tillman JB, Wardlaw D, et al. A randomized trial of balloon kyphoplasty and nonsurgical management for treating acute vertebral compression fractures: vertebral body kyphosis correction and surgical parameters. Spine (Phila Pa 1976) 2013;38:971–83.
101. Stienen MN, Joswig H, Smoll NR, Corniola MV, Schaller K, Hildebrandt G, et al. Influence of body mass index on subjective and objective measures of pain, functional impairment, and health-related quality of life in lumbar degenerative disc disease. World Neurosurg 2016;96:570–7.
102. Stienen MN, Smoll NR, Joswig H, Snagowski J, Corniola MV, Schaller K, et al. Influence of the mental health status on a new measure of objective functional impairment in lumbar degenerative disc disease. Spine J 2017;17:807–13.
103. Stienen MN, Maldaner N, Sosnova M, Joswig H, Corniola MV, Regli L, et al. Lower extremity motor deficits are underappreciated in patient-reported outcome measures: added value of objective outcome measures. Neurospine 2020;17:270–80.
104. Tenhula J, Lenke LG, Bridwell KH, Gupta P, Riew D. Prospective functional evaluation of the surgical treatment of neurogenic claudication in patients with lumbar spinal stenosis. J Spinal Disord 2000;13:276–82.
105. Yukawa Y, Lenke LG, Tenhula J, Bridwell KH, Riew KD, Blanke K. A comprehensive study of patients with surgically treated lumbar spinal stenosis with neurogenic claudication. J Bone Joint Surg Am 2002;84:1954–9.
106. Barz T, Melloh M, Staub L, Roeder C, Lange J, Smiszek FG, et al. The diagnostic value of a treadmill test in predicting lumbar spinal stenosis. Eur Spine J 2008;17:686–90.
107. Herno A, Airaksinen O, Saari T. Computed tomography after laminectomy for lumbar spinal stenosis: patients' pain patterns, walking capacity, and subjective disability had no correlation with computed tomography findings. Spine (Phila Pa 1976) 1994;19:1975–8.

108. Herno A, Airaksinen O, Saari T, Pitkänen M, Manninen H, Suomalainen O. Computed tomography findings 4 years after surgical management of lumbar spinal stenosis: no correlation with clinical outcome. Spine (Phila Pa 1976) 1999;24:2234–9.
109. Koc Z, Ozcakir S, Sivrioglu K, Gurbet A, Kucukoglu S. Effectiveness of physical therapy and epidural steroid injections in lumbar spinal stenosis. Spine (Phila Pa 1976) 2009;34:985–9.
110. Malmivaara A, Slätis P, Heliövaara M, Sainio P, Kinnunen H, Kankare J, et al. Surgical or nonoperative treatment for lumbar spinal stenosis? A randomized controlled trial. Spine (Phila Pa 1976) 2007;32:1–8.
111. Geisser ME, Haig AJ, Tong HC, Yamakawa KS, Quint DJ, Hoff JT, et al. Spinal canal size and clinical symptoms among persons diagnosed with lumbar spinal stenosis. Clin J Pain 2007;23:780–5.
112. Rainville J, Childs LA, Peña EB, Suri P, Limke JC, Jouve C, et al. Quantification of walking ability in subjects with neurogenic claudication from lumbar spinal stenosis: a comparative study. Spine J 2012;12:101–9.
113. Whitman JM, Flynn TW, Childs JD, Wainner RS, Gill HE, Ryder MG, et al. A comparison between two physical therapy treatment programs for patients with lumbar spinal stenosis: a randomized clinical trial. Spine (Phila Pa 1976) 2006;31:2541–9.
114. Zeifang F, Schiltenwolf M, Abel R, Moradi B. Gait analysis does not correlate with clinical and MR imaging parameters in patients with symptomatic lumbar spinal stenosis. BMC Musculoskelet Disord 2008;9:89.
115. Markman JD, Gewandter JS, Frazer ME, Murray NM, Rast SA, McDermott MP, et al. A randomized, double-blind, placebo-controlled crossover trial of oxymorphone hydrochloride and propoxyphene/acetaminophen combination for the treatment of neurogenic claudication associated with lumbar spinal stenosis. Spine (Phila Pa 1976) 2015;40:684–91.
116. Henchoz Y, de Goumoëns P, Norberg M, Paillex R, So AK. Role of physical exercise in low back pain rehabilitation: a randomized controlled trial of a three-month exercise program in patients who have completed multidisciplinary rehabilitation. Spine (Phila Pa 1976) 2010;35:1192–9.
117. Deen HG, Zimmerman RS, Lyons MK, McPhee MC, Verheijde JL, Lemens SM. Use of the exercise treadmill to measure baseline functional status and surgical outcome in patients with severe lumbar spinal stenosis. Spine (Phila Pa 1976) 1998;23:244–8.
118. Fritz JM, Erhard RE, Delitto A, Welch WC, Nowakowski PE. Preliminary results of the use of a two-stage treadmill test as a clinical diagnostic tool in the differential diagnosis of lumbar spinal stenosis. J Spinal Disord 1997;10:410–6.
119. Tomkins CC, Battié MC, Rogers T, Jiang H, Petersen S. A criterion measure of walking capacity in lumbar spinal stenosis and its comparison with a treadmill protocol. Spine (Phila Pa 1976) 2009;34:2444–9.
120. Deen HG, Zimmerman RS, Lyons MK, McPhee MC, Verheijde JL, Lemens SM. Test-retest reproducibility of the exercise treadmill examination in lumbar spinal stenosis. Mayo Clin Proc 2000;75:1002–7.
121. Whitehurst M, Brown LE, Eidelson SG, D'angelo A. Functional mobility performance in an elderly population with lumbar spinal stenosis. Arch Phys Med Rehabil 2001;82:464–7.
122. Papavero L, Thiel M, Fritzsche E, Kunze C, Westphal M, Kothe R. Lumbar spinal stenosis: prognostic factors for bilateral microsurgical decompression using a unilateral approach. Neurosurgery 2009;65:182–7.
123. Yaşar B, Simşek S, Er U, Yiğitkanli K, Ekşioğlu E, Altuğ T, et al. Functional and clinical evaluation for the surgical treatment of degenerative stenosis of the lumbar spinal canal. J Neurosurg Spine 2009;11:347–52.
124. Drury T, Ames SE, Costi K, Beynnon B, Hall J. Degenerative spondylolisthesis in patients with neurogenic claudication effects functional performance and self-reported quality of life. Spine (Phila Pa 1976) 2009;34:2812–7.
125. Azad TD, Kalani M, Wolf T, Kearney A, Lee Y, Flannery L, et al. Building an electronic health record integrated quality of life outcomes registry for spine surgery. J Neurosurg Spine 2016;24:176–85.
126. McCormick JD, Werner BC, Shimer AL. Patient-reported outcome measures in spine surgery. J Am Acad Orthop Surg 2013;21:99–107.
127. Resnik L, Dobrzykowski E. Guide to outcomes measurement for patients with low back pain syndromes. J Orthop Sports Phys Ther 2003;33:307–16.
128. Maldaner N, Desai A, Gautschi OP, Regli L, Ratliff JK, Park J, et al. Improving the patient-physician relationship in the digital era: transformation from subjective questionnaires into objective real-time and patient-specific data reporting tools. Neurospine 2019;16:712–4.
129. Stienen MN, Gautschi OP, Staartjes VE, Maldaner N, Sosnova M, Ho AL, et al. Reliability of the 6-minute walking test smartphone application. J Neurosurg Spine 2019;13:1–8.
130. Maldaner N, Tomkins-Lane C, Desai A, Zygourakis CC, Weyerbrock A, Gautschi O, et al. Digital transformation in spine research and outcome assessment. Spine J 2020;20:310–1.

Table 1. Practical applications*

Measure	Number of Items	Content/Domains	Method of Administration	Recall Period	Response Format	Range of Scores	Score Interpretation	Availability of Normative Data	Cross-Cultural Validation
ODI	10	One item on pain and nine items on activities of daily living	Self-completed questionnaire by patient	Today	6-point ordinal scale	0 (no pain/ disability)-50 points (maximal pain/disability) or 0% (no pain/ disability)-100% (maximal pain/ disability)	0 (no disability)-100 (maximum disability); 0-20: minimal disability; 20-40: moderate disability; 40-60: severe disability; 60-80: housebound; 80-100: bedbound	Few data exist. Normative scores range from 8-10 according to study population	Excellent
RMDQ	24	Questions on daily physical activities and functions (housework, sleeping, mobility, dressing, getting help, appetite, irritability, and pain severity)	Self-completed by patient	Past 24h	Each item is given a score of either a 1 (agree with statement) or 0 (disagree with statement).	0 (no disability) - 24 (maximal disability)	Higher scores indicate higher disability.	None are known to the author. It is estimated that an RMDQ score of <2 distinguished between patients who considered themselves asymptomatic.	Excellent
SSM	18	Seven questions on symptom severity (SSM symptoms), five questions on physical function (SSM function), and six questions on satisfaction (SSM satisfaction)	Self-completed questionnaire by patient	Past month	3-5–point ordinal scale	SSM function: 1-4 points; SSM symptoms: 1-5 points	Higher scores indicate higher disability/more severe symptoms/ less satisfaction.	None are known to the authors.	Excellent
COMI Back	7	Pain intensity, back-related function, symptom-specific well-being, general quality of life, and social disability/ work disability	Self-completed questionnaire by patient	Past week	Pain intensity is measure on a 0-10 numeric rating scale. Other items are measured on a 5-point ordinal scale.	0-10 points	Higher scores indicate higher disability.	None	Excellent

(Continued)

Table 1. (*Cont'd*)

Measure	Number of Items	Content/Domains	Method of Administration	Recall Period	Response Format	Range of Scores	Score Interpretation	Availability of Normative Data	Cross-Cultural Validation
NDI	10	Pain intensity, personal care, lifting, work, headache, concentration, sleeping, driving, reading, and recreation	Self-completed questionnaire by patient	Current status	6-point ordinal scale	0 (no pain/ disability)-50 points (maximal pain/disability) or 0% (no pain/ disability)-100% (maximal pain/ disability)	0-4 points (0%-8%): no disability; 5-14 points (10%-28%): mild disability; 15-24 points (30%-48%): moderate disability; 25-34 points (50%-68%): severe disability; >35 points (>70%): complete disability	Few data exist. A Japanese study with 1200 participants showed a mean NDI score of 6.98	Excellent
TUG test	–	The patient is asked to stand up, walk back and forth 3 m, and sit down again as fast as possible, while time is measured.	Examiner-based task. No special equipment required (besides a stopwatch, chair, and 3-m walking space). A free smartphone app ("TUG app") is available.	–	–	Expressed as raw TUG test time (in seconds) and as standardized *z* score/T score adjusted for age and sex and based on normal population reference values	Patients with raw TUG test results >12 s, *z* scores >2.3, and T scores > 123.0 are considered to have OFI. OFI can be further stratified according to severity index into no, mild, moderate, and severe OFI.	Yes	No translation needed; the TUG test has been used in different countries.
MTT	–	The patient is instructed to walk on a calibrated treadmill, usually for 15-30 min.	Observer-led task; special equipment and trained personal are required.	–	–	Expressed in raw test times of the time to first symptoms, total ambulation time and distance, and maximum walking speed.	No standardized T or *z* scores; different walking protocols impede comparison between studies.	None	No translation needed; the MTT has been used in different countries.

* COMI = Core Outcome Measures Index; MTT = motorized treadmill test; NDI = Neck Disability Index; ODI = Oswestry Disability Index; RMDQ = Roland-Morris Disability Questionnaire; SSM = Spinal Stenosis Measure; TUG = Timed Up-and-Go.

Table 2 . Psychometrics*

Measure	Floor and Ceiling Effects	Reliability	Validity	Responsiveness	MCIDs	Generalizability	Used in RCTs
ODI	Unclear	Good ICC and internal consistency	Adequate construct and content validity; lacks component of generic HRQOL	Good	Cutoff point for MCID is 10 points or 30% score improvement	Excellent	Yes; most widely used PROM for lumbar DDD
RMDQ	No improvement detected for scores lower than 4; no decline in score equal to 20.	Internal consistency (Cronbach's α = 0.84-0.93) and ICC are good. MDC and SEM are known but are influenced by factors like time intervals and methods used.	Acceptable; correlates well with SF-36, ODI, Sickness Impact Profile, Quebec Back Scale, and VAS/NRS	Moderate to large responsiveness (response mean = 0.78-0.84 for improvement)	MCID = 2-5 points and varies depending on the patient's initial score. A 30% change from baseline was proposed as MCID.	Can be used in acute and subacute patients and patients with chronic LBP	Yes, extensively
SSM	Unclear	Good ICC	Good construct validity; high correlation with other questionnaires, including SF-36 and Oxford Spinal Stenosis Score	SRM = 0.96-1.43 for the subscales and SRM = 1.50 for the average scale	MCID = 0.5-2 for SSM function and MCID = 0.48 for SSM symptoms	Specific to patients with LSS; less commonly used for general LBP population	Yes
COMI Back	Some items are reported to have floor or ceiling effects of 20%-50%.	Good ICC and internal consistency	Moderate to high correlation with other questionnaires (ODI and SF-36)	SRM =0.62-1.44; AUC analysis showed good discriminative abilities between treatment results.	MCID = 2-3 points	Excellent	Yes
NDI	Baseline scores in the lower 10th and 90th percentiles are subject to significant floor and ceiling effects	ICC > 0.90	High correlation with other neck-specific questionnaires; good internal consistency	SRM = 0.60-0.95	MCIDs range from 5/50 to 19/50; an average change of 7 points can be regarded as clinically meaningful	Excellent	Yes; most widely used PROM for neck pain
TUG test	Likely to have a floor but not a ceiling effect	Excellent intra- and interrater reliability; SEM = 0.21-0.23 s	Significant moderate correlation with various metrics, including VAS back/leg, ODI, SF-12, and EQ-5D; likely to measure a different dimension of patient's functional status compared with PROMs	SRM = 0.51	The MCID = 3.4 s in raw TUG test time for patients with lumbar DDD	Excellent	Yes; most commonly used objective functional outcome measure for lumbar DDD
MTT	Unclear	High to excellent ICC	Good convergent validity with other walking tests and self-reported symptoms; lower correlation with PROMs (eg, ODI)	Good responsiveness is reported; however, no SRMs are available.	Unclear	Relatively specific to LSS patients; less commonly used for general LBP population	Yes

* AUC = area under the curve; COMI = Core Outcome Measures Index; DDD = degenerative disk disease; EQ-5D = EuroQol Five Dimension Questionnaire; ICC = intraclass correlation coefficient; LBP = low back pain; LSS = lumbar spinal stenosis; MCID = minimum clinically important difference; MDC = minimal detectable change; MTT = motorized treadmill test; NDI = Neck Disability Index; NRS = numeric rating scale; ODI = Oswestry Disability Index; PROM = patient-reported outcome measure; RMDQ = Roland-Morris Disability Questionnaire; SF-12 = Short Form 12; SF-36 = Short Form 36; SRM = standardized response mean; SSM = Spinal Stenosis Measure; TUG = Timed Up-and-Go; VAS = visual analog scale.